Using Generative Adversarial Networks for Network Intrusion Detection

XuDong Li[1,2], Di Lin[1,2(✉)], Yu Tang[2], Weiwei Wu[2], Zijian Li[2], and Bo Chen[1,2]

[1] Intelligent Terminal Key Laboratory of Sichuan Province, Yibin, China
lindi@uestc.edu.cn
[2] University of Electronic Science and Technology of China, Chengdu, China

Abstract. The network intrusion detection system is an essential guarantee for network security. Most research on network intrusion detection systems focuses on using supervised learning algorithms, which require a large amount of labeled data for training. However, the work of labeling data is complex and cannot exhaustively include all types of network intrusion. Therefore, in this study, we develop a model that only requires normal data in the training phase, and it can distinguish between normal data and abnormal data in the test phase. This model is implemented by using a generative confrontation network. Experimental results show that, on the CIC-IDS-2017 dataset, our model has an accuracy of 97%, which is dramatically higher than the basic autoencoder, which is one of the most widely used algorithms in the network intrusion detection.

Keywords: Generative adversarial network · Network intrusion detection · Network security

1 Introduction

With the increasing complexity of modern networks, the growing popularity of network use, the growing diversification of network attacks, and the rapid development of the Internet, more and more devices are connected to the network, and there are significant challenges to network security. The demand for network intrusion detection systems is therefore increasing. The biggest problem facing the current network intrusion detection system is the lack of awareness of strange events, the lack of detection of unknown risks such as zero-day attacks, and the low detection rate of low-frequency attacks such as worm attacks [1]. If modern network system security still relies on manual detection by administrators, identification and processing are inefficient. So the best solution is to let the machine learn the ability to analyze network data and detect any suspicious or abnormal behavior. Relying on the powerful ability of deep neural networks to extract data features [2] automatically can be achieved.

According to different methods of detecting anomalies, network intrusion detection systems can be divided into signature-based methods and anomaly detection-based methods [1, 3]. Both methods have their own advantages and disadvantages. The

S. Shi et al. (Eds.): 6GN 2021, LNICST 439, pp. 61–68, 2022.
https://doi.org/10.1007/978-3-031-04245-4_6

signature-based method describes the known attacks in detail. It can efficiently and accurately detect various known attacks. However, such methods are not capable of dealing with unknown threats. For example, zero-day attacks cannot be handled. Such attacks often bring more significant harm. The method based on anomaly detection is suitable for responding to unknown threats and can detect unknown or new types of attacks. Anomaly-based network intrusion detection systems can be divided into two categories: network intrusion detection systems based on supervised learning and network intrusion detection systems based on unsupervised learning [4]. Current researches mainly focus on using supervised learning to build intrusion detection systems. Still, the problem is that a large amount of labeled data is needed, and it is almost impossible to obtain a data set that includes all types of attacks because network attack methods are endless. Therefore, this paper uses a generative confrontation network to implement a model that only needs normal data in the training phase. In this process, the model captures the distribution of normal data. In the test phase, the model can distinguish between normal data and abnormal data. To realize the classification task, the data that does not conform to the known distribution is judged as abnormal.

This paper confirms the feasibility of using a Generative Adversarial Network in network intrusion detection. In Sect. 2, the previous research in the field of IDS is discussed. Section 3 first introduces the CIC-IDS-2017 data set used in the experiment. Then the proposed model and corresponding detection framework are presented. In Sect. 4, the evaluation indicators used are explained. Then the performance of the proposed model is evaluated, and the results are discussed. Finally, Sect. 5 discusses the significance and limitations of this study and makes a reasonable outlook for future development.

2 Relate Work

According to different technologies, the current research on network intrusion detection is carried out from the following three directions.

A rule-based approach. This type of method is designed based on the characteristics of known attacks and has sound effects on known attacks but has obvious shortcomings in dealing with unknown threats. A simple rule-based system network intrusion detection system can not meet current industry needs.

Based on traditional machine learning methods. In [5], the author uses a random forest plus XGBoost method to implement a network intrusion detection model and introduces a cost-sensitive function to improve the detection rate of a small sample category. In [6], a 10-fold cross-validation decision tree method was used to study network intrusion detection on a data set containing 22 features (feature selection was made on the NSL-KDD data set). The study in [7] found that random forest performs best in this type of problem. Traditional machine learning algorithms have achieved certain results in network intrusion detection tasks, but they also have limitations. For example, standard machine learning algorithms require manual feature engineering of data to construct sample features.

Based on deep learning methods. Such methods are the best prospects. In response to known attacks, it can make automatic feature selection on a high accuracy rate and low false alarm rate. If the appropriate network architecture design, it also has a role in

dealing with unknown threats. For example, in [8], the combined Sparse Auto-Encoder and soft-max model are used to do network intrusion detection. It uses Sparse Auto-Encoder to perform feature learning on the unlabeled data set to obtain the hidden layer encoding. Then take the hidden layer obtained from the previous training step and add soft-max to classify it on the labeled data set to complete the detection. In [9], an asymmetric stacked autoencoder (S-NDAE) plus random forest is used for network intrusion detection. The autoencoder is used for feature extraction, and the random forest is used for classification. In [10], the CNN network is used for network intrusion detection. The detection is completed by increasing the convolution kernel to map the original features to the high-dimensional space to enhance the feature learning ability. In [11], the author uses an improved convolutional neural network to implement a network intrusion detection model. This model uses a cross-layer aggregation network design method, which differs from the traditional convolution-pooling-full connection structure. Experimental results show that this model has achieved good results. In [12], LSTM network architecture is used for network intrusion detection. It uses an autoencoder to extract data features. Using the timing-related characteristics of network data and LSTM network for detection, this method has a specific effect in dealing with unknown threats. GAN network also has specific applications in the field of intrusion detection. For example, in [13], the author used the GAN-PSO-ELM model to conduct network intrusion detection research. It uses GAN to expand the minority samples, uses PSO to optimize the input weight and hidden layer bias of ELM, and finally builds the model. Experiments on the NSL-KDD data set have achieved good results.

3 Model Design

3.1 Dataset

The CIC-IDS-2017 data set used in this paper was published by the Canadian Institute of Cyber Security (CIC). This data set collects 5 days of data, which contains network data under normal conditions and network data under the latest common attacks. It builds an abstract behavior of 25 users based on HTTP, HTTPS, FTP, SSH, and email protocols, which simulates the real-world situation to the greatest extent. It also includes the results of network traffic analysis using CICFlowMeter, which is based on timestamps, source and destination IP, source and destination ports, protocol, and attack marking traffic [14]. Current network intrusion detection research recommends using this data set compared with data sets such as KDD and NSL-KDD.

3.2 Data Processing

Each record of the CIC-IDS-2017 data set contains 78 features, all of which are numerical. The data processing for this includes the following steps:

Step 1: Eliminate data with missing features in the data set.
Step 2: To better convert the data into an image, three all-zero features are added. Each record is composed of 91 features, and the last three features of each record are all 0.

Step 3: Normalize the data to eliminate the influence of extreme data on the model, which helps deep learning model training. The normalization formula is as follows:

$$x = \frac{x - Min}{Max - Min} \tag{1}$$

3.3 Network Architecture Design

In this section, the proposed GAN model will be described. Its generator and discriminator both use convolutional neural networks.

1) Design discriminator

The structure of the discriminator is shown in Fig. 1. Conv2D, LeakyReLu, and Dropout are one layer, a total of three layers are superimposed, and finally, after the Flatten operation, a Dense output result is connected. In the last layer, we did not use the sigmoid function. The discriminator receives a $9 \times 9 \times 1$ size picture and outputs a score for this picture. The scoring intervals for normal data and abnormal data are different. Use this feature to identify abnormal and normal states in the network.

2) Design generator

The structure of the generator is shown in Fig. 2. The generator receives random noise data and expands it to the same size as the real data for input into the discriminator. The generator also uses the leak-relu activation function, and to get better results, we also added batch normalization. The generator is to help the discriminator to train. The generator is useless when the training is complete.

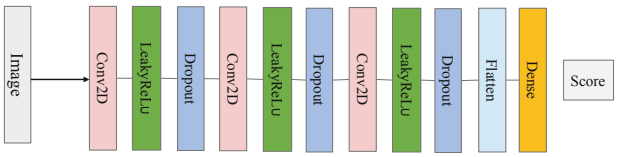

Fig. 1. The discriminator structure of the proposed GAN model

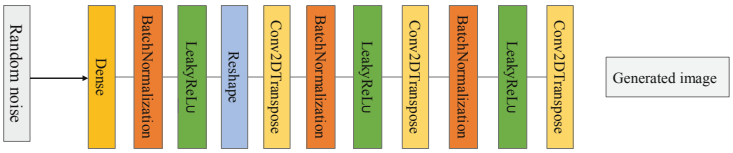

Fig. 2. The generator structure of the proposed GAN model

3.4 Intrusion Detection Framework

The generator's goal in GAN is to generate fake data that is as close to the real data as possible. The purpose of the discriminator is to distinguish whether the input data comes

from real data or fake data generated by the generator. After the training process, the discriminator has captured the distribution of real data, the distribution of normal data. Its score for normal data output will stabilize within a range, and its score for abnormal data output that has never been seen will deviate from this range. The study found that the score distribution of the normal data output by the discriminator is similar to the normal distribution. Therefore, the "3σ" principle of a normal distribution is used to determine the division of threshold points in the normal data range. The mean value μ of the data plus or minus three times the standard deviation σ to determine. The threshold division formula is as follows.

$$\text{threshold} = (\mu - 3\sigma, \mu + 3\sigma) \tag{2}$$

The steps of using GAN to detect abnormal data in the network are shown in Fig. 3.

1) To process the original data, the specific steps are as described in Sect. 3.2. After processing, 70% of the normal data is used for training GAN and 30% for testing the model. 100% of abnormal data is used to test the model. It should be noted that the model has never seen 30% of normal data and all abnormal data during the training process.
2) Only use normal data to train GAN. After the training is completed, the generator can generate normal data close to the real, and the discriminator captures normal data distribution.
3) Performance measurement. Take the unused normal data and mix it with the abnormal data. Use the combined data to test the performance of the model.

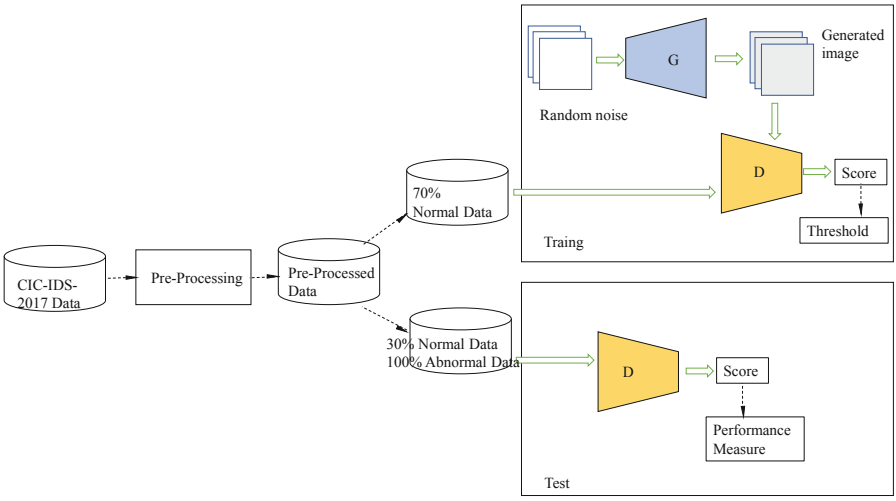

Fig. 3. The framework of the network intrusion detection process using the proposed GAN model includes the training part and the testing part. The threshold division of the training part is carried out according to formula 2.

4 Experiment

4.1 Evaluation Index

There are four evaluation indicators used in this study.

Precision(P): The percentage of intrusions that are correctly judged as intrusions.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

Recall(R): judicious invasion percentage of all intrusion traffic.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

Accuracy(A): The ratio of the total number of flows that are correctly judged.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{5}$$

F value: the result of weighted and averaged precision and recall.

$$\text{F} = \frac{2PR}{P + R} \tag{6}$$

Among them, TP is the number of samples of attack behaviors that are correctly classified;

TN is the number of samples of normal behaviors that are correctly classified;
FP is the number of samples of normal behaviors that are misclassified;
FN is the number of examples of misclassified attack behaviors.

4.2 Experimental Results

In the CIC-IDS-2017 data set, the data is divided into normal data and abnormal data (that is, data at the time of the attack). Take 70% of the normal data to train the GAN model. When the training is completed, take the discriminator in the GAN model to achieve the detection task. The visualization of the score obtained by 70% of the normal data through the trained discriminator is shown in Fig. 4. It can be found that it is similar to the normal distribution. Unfortunately, we found that it does not entirely conform to the normal distribution through testing. This is also the reason for the design of the threshold division formula. After that, we take the same amount of abnormal and normal data for experiments, use them to obtain scores through the trained discriminator, and then visualize the results. As shown in Fig. 5, we find that normal data scores and abnormal data scores are gathered in different clusters. This is one of the reasons why our model can work.

The performance of our proposed model was tested on the CIC-IDS-2017 data set and compared with the basic autoencoder. The results are shown in Table 1. According to Table 1, the accuracy of our proposed model is 97.64%, the precision is 99.43%, the recall rate is 95.82%, and the F value is 0.97. All indicators are better than the results obtained by the basic autoencoder. At the same time, these good experimental results also reflect that our proposed model can accurately and comprehensively detect normal data and abnormal data.

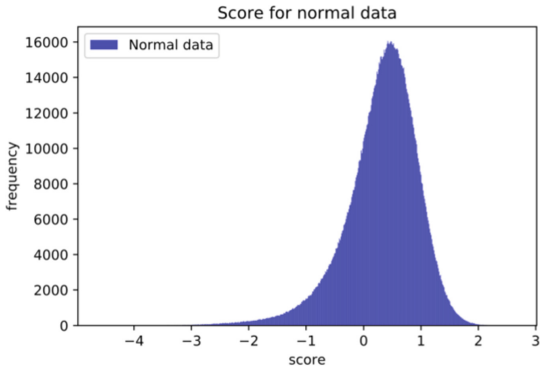

Fig. 4. The normal data used for training is visualized by the score after the trained discriminator

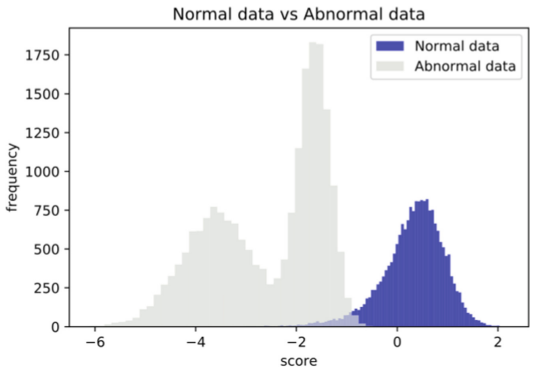

Fig. 5. The normal data and abnormal data used in the test are visualized by the scores of the trained discriminator.

Table 1. Performance comparison between proposed GAN model and basic autoencoder

Model	Accuracy(%)	Precision(%)	Recall(%)	F(%)
GAN discriminator	97.64	99.43	95.82	97.59
Basic autoencoder	84.0	79.39	91.85	85.17

5 Conclusion

In this paper, we propose a Generative Adversarial Network based network intrusion detection algorithm. Specifically, a framework of Generative Adversarial Network is employed to complete intrusion detection and conducts experiments. Experimental results show that this method has a certain degree of stability, and it also outperforms the current methods, e.g., a basic autoencoder. The method proposed in this paper for network intrusion detection can effectively solve the problem of detecting unknown attacks, which is a challenging problem.

Acknowledgment. Partially Funded by Science and Technology Program of Sichuan Province (2021YFG0330), partially funded by Grant SCITLAB-0001 of Intelligent Terminal Key Laboratory of SiChuan Province,and partially Funded by Fundamental Research Funds for the Central Universities (ZYGX2019J076).

References

1. Nassar, M., et al.: Network intrusion detection, literature review and some techniques comparision. In: International Computer Engineering Conference (2019)
2. Lecun, Y., Bengio, Y., Hinton, G.E.: Deep learning. Nature **521**(7553), 436–444 (2015)
3. Vinayakumar, R., et al.: Applying convolutional neural network for network intrusion detection. In: Advances in Computing and Communications, pp. 1222–1228 (2017)
4. Gogoi, P., Borah, B., Bhattacharyya, D.K.: Anomaly detection analysis of intrusion data using supervised & unsupervised approach. J. Converg. Inf. Technol. **5**(1), 95–110 (2010)
5. Chen, Z., Lyu, N.: Network intrusion detection model based on random forest and XGBoost. J. Signal Process. Syst. **36**(7), 1055–1064 (2020)
6. Chae, H.S., Jo, B.O., Choi, S.H., Park, T.K.: Feature selection for intrusion detection using NSL-KDD. In: Recent Advances in Computer Science, pp. 184–187 (2013)
7. Thaseen, S., Kumar, C.A.: An analysis of supervised tree based classiers for intrusionv detection system. In: Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on, pp. 294–299, IEEE (2013)
8. Javaid, A., et al.: A deep learning approach for network intrusion detection system. In: Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS) (2016)
9. Shone, N., et al.: A deep learning approach to network intrusion detection. IEEE Trans. Emerg. Top. Comput. Intell. **2**(1), 41–50 (2018)
10. Khan, R.U., et al.: An improved convolutional neural network model for intrusion detection in networks. In: 2019 Cybersecurity and Cyberforensics Conference (CCC). IEEE (2019)
11. Yang, H., Wang, F.: Network intrusion detection model based on improved convolutional neural network. J. Comput. Appl. **39**(9), 2604–2610 (2019)
12. Mirza, A.H., Selin, C.: Computer network intrusion detection using sequential LSTM neural networks autoencoders. In: 2018 26th Signal Processing and Communications Applications Conference (SIU). IEEE (2018)
13. Yang, Y., Song, R., Zhou, Z.: Network intrusion detection method based on GAN-PSO-ELM. Comput. Eng. Appl. **56**(12), 66–72 (2020)
14. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: International Conference on Information Systems Security & Privacy (2018)